International Journal of Approximate Reasoning 53 (2012) 541–557

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Bayesian rough set model: A further investigation

Hongyun Zhang, Jie Zhou *, Duoqian Miao, Can Gao

Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China

ARTICLE INFO

Article history:
Received 1 March 2009
Received in revised form 24 June 2010
Accepted 14 December 2011
Available online 22 December 2011

Keywords:
Bayesian rough set model
Binary decision
Multi-decision
Discernibility matrix
Attribute reduction

ABSTRACT

Bayesian rough set model (BRSM), as the hybrid development between rough set theory and Bayesian reasoning, can deal with many practical problems which could not be effectively handled by original rough set model. In this paper, the equivalence between two kinds of current attribute reduction models in BRSM for binary decision problems is proved. Furthermore, binary decision problems are extended to multi-decision problems in BRSM. Some monotonic measures of approximation quality for multi-decision problems are presented, with which attribute reduction models for multi-decision problems can be suitably constructed. What is more, the discernibility matrices associated with attribute reduction for binary decision and multi-decision problems are proposed, respectively. Based on them, the approaches to knowledge reduction in BRSM can be obtained which corresponds well to the original rough set methodology.

1. Introduction

Rough set theory [1,2] is aimed at data analysis problems involving uncertain, imprecise or incomplete information. Since it was introduced by Pawlak in 1982, it has been successfully used in many research fields, such as pattern recognition, machine learning, knowledge acquisition, economic forecasting and data mining [3–7]. Knowledge classification is a fundamental problem in rough set theory. In Pawlak's rough set model, the degree of set overlap was not considered, namely, the classification must be totally correct or certain. Therefore, original rough set model cannot effectively deal with data sets which have noisy data and latent useful knowledge in the boundary region may not be fully captured. In order to overcome the limitations, some extended rough set models have been put forward which combine with other available soft computing technologies, such as statistical rough set model [8], decision-theoretic rough set model [26,27], fuzzy rough set model [9,10], covering rough set model [11], tolerance rough set model [12], dominance-based rough set model [13,14] and others.

Many researchers were motivated to investigate probabilistic approaches to rough set theory [15–17,26,33,39,47]. Variable precision rough set model (VPRSM) [17] is one of the most important extensions. In the model, standard inclusion relation is extended to majority inclusion relation, and the novel notion can be able to allow for some degree of misclassification in the largely correct classification. The strict functional or dependent relations between attributes will be softened. As a result, more general association decision rules including deterministic and probabilistic ones can be obtained in VPRSM. Subsequently, Ziarko et al put forward an asymmetric variable precision rough set model (AVPRSM) [18], and the model becomes more general and flexible. Variable precision rough set models, symmetric or asymmetric, involve some parameters, β or $\{l, u\}$. Different parameters will result in different models, and the extracted decision rule sets may be distinct. In the applications, it is not clear how to find out the optimal parameters and their values are often selected based on the decision makers' previous knowledge of the domain and their intuition or the proposed criteria [19–23].

The connections between rough sets and Bayes' theory were analyzed by Pawlak [24,25]. Rough set theory offers a new view on Bayes' theory, and any decision data set in rough set theory will satisfy the total probability theorem and Bayes'

* Corresponding author.
E-mail addresses: zhanghongyun583@sina.com (H.Y. Zhang), jie_jpu@163.com (J. Zhou), miaoduoqian@163.com (D. Miao), 2005gaocan@163.com (C. Gao).

0888-613X/$ - see front matter
http://dx.doi.org/10.1016/j.ijar.2011.12.006

theorem. Based on Bayesian decision procedure with minimum risk, Yao [26–29] put forward a new model called decision-theoretic rough set model(DTRSM) which brings new insights into the probabilistic approaches to rough set theory. DTRSM provides a general framework for comparing and synthesizing probabilistic rough set approximations. It not only has good theoretical foundation, but also possesses reasonable semantic interpretation [30,31]. The Pawlak's rough set model, VPRSM and AVPRSM can be directly derived from DTRSM under relevant loss functions. If the practical decision problems involve cost or risk environments, the DTRSM will be more beneficial for decision making compared with original rough set model. Moreover, VPRSM and AVPRSM can be considered as an intermediate step when using the decision theoretic approach for rough analysis [32].

According to Bayesian reasoning, Slezak and Ziarko [33] presented Bayesian rough set model (BRSM), in which the parameters that control the approximation regions in VPRSM are determined by the prior probability of occurrence of the target event under consideration. As the hybrid development, BRSM is reasonable for handling some practical domains in which the prior probability of the assumption will be affected by adding some new evidences, such as medical diagnosis, fault detection, economy forecasting and so on. Since the prior probability can be estimated from data set itself, BRSM is more objective when compared with the parametric versions of variable precision rough set model.

A more general parametric modification of BRSM, called variable precision Bayesian rough set model(VPBRSM), was further proposed by Slezak [34,35] which allows single parameter-controlled degree of ε-precision in the approximation region definition. VPBRSM is more applicable to practical data analysis problems where small deviations from prior probability are likely to occur due to noise or measurement inaccuracy [36–38]. Based on the Bayes factor and the inverse probabilities, Slezak also introduced a parameterized extension of rough set model, called rough Bayesian model(RBM) [39,40] which can be utilized very well if the prior and posterior probabilities derivable from data or background knowledge are not reliable. Under associated parameters, the Pawlak's rough set model, VPRSM and BRSM can also be derivable from RBM.

Attribute reduction is one of the most fundamental and important notions in rough set theory. A reduct is a minimal subset of attributes that preserves a certain classification property as provided by the entire set of attributes. The monotonic property of approximation quality along with reducing attributes in Pawlak's rough set model will not be satisfied in the probabilistic rough set models. This could bring some anomalies to the procedures of attribute reduction if Pawlak's classical reduct definition will be directly applied [31,41,42]. How to establish a reasonable objective function for attribute reduction is a pivotal problem that needs to be circumvented in the probabilistic approaches to rough set theory.

In this paper, we concentrate on data analysis problems including binary decision and multi-decision cases with non-parametric Bayesian rough set model. For binary decision problems, Slezak [43] used global relative gain function and Ziarko [44,47] used normalized expected absolute gain function to construct attribute reduction model in BRSM, respectively. After being analyzed critically, it is found that these two attribute reduction models are equivalent in essence. Furthermore, we extend attribute reduction model from binary decision problems to multi-decision problems in BRSM. Some monotonic approximation quality measures are presented for multi-decision problems, and their relationships are also discussed. Therefore, attribute reduction model for multi-decision problems can be constructed based on these monotonic quantitative measures. Skowron [45] has proved that reducts are in one-to-one correspondence to the prime implicants of the associated discernibility function in a given decision table. According to this property, discernibility matrices for binary decision and multi-decision problems are introduced under corresponding attribute reduction models in BRSM, respectively. It is also illustrated that the attribute reduction model for binary decision problems is a special case of multi-decision problems, and then an uniform formulation of attribute reduction in BRSM can be described.

The remainder of this paper is organized as follows. Section 2 reviews some basic notions of rough set theory. Some probabilistic approaches to rough set model are introduced in Section 3 and comparative properties among them will also be investigated. Section 4 discusses the equivalence between two kinds of current attribute reduction models in BRSM for binary decision problems. Section 5 presents some approximation quality measures as well as their relationships for multi-decision problems. The attribute reduction model in BRSM for multi-decision problems is constructed in Section 6. In Section 7, main conclusions are covered.

2. Preliminaries

In this section, some basic notions will be briefly reviewed. More detailed descriptions can be found in [1,2,29,44,46].

2.1. Information system

The object sets discussed in rough set theory are presented as information systems. An information system is the tuple: $IS = (U, A, V, \rho)$. U is a universe of discourse. A is a set of attributes that describe the objects in the universe U. $V = \cup_{a\in A} V_a$ is the union of attribute value domains, V_a is a nonempty set of values for attribute $a \in A$. $\rho : U \times A \to V$ is an information function, $\rho(x, a)$ means that object x has the value on attribute a.

$\forall B \subseteq A$ defines an equivalence relation, referred to as an indiscernibility relation and presented as: $IND(B) = \{(x, y) | \forall b \in B \to \rho(x, b) = \rho(y, b), x, y \in U\}$. With equivalence relation $IND(B)$, universe U can be partitioned into a collection of equivalence classes $U/IND(B)$, denoted as U/B briefly. Each $E \in U/B$ is called an elementary set with respect to B. Obviously, the partition U/A has the finest granularity and the partition $U/\emptyset$ has the coarsest granularity.

Strictly speaking, universe U should be infinite in the real world in most cases and we cannot get the whole objects in it. Practically, rough set theory can only deal with a finite nonempty sample set of the whole universe $\tilde{U} \subseteq U$. Yao [30], Ziarko [44], Slezak [39] and others have systematically discussed the probabilistic approach to rough set theory. Statistically, probabilistic knowledge reflects the relative occurrence frequencies of any subsets (or called events) in universe U. Then the knowledge in rough set theory can be considered as random variables which are represented by the probability functions defined on the σ algebra of measurable subsets of universe U.

For subset $X \subseteq U$, if $X = U$ or $X = \emptyset$, then X is called a trivial subset in U. In the applications, trivial subsets are not interested in. Because the probabilities of their occurrence are certain. The prior probability of event X is denoted as $P(X)$ and its condition probability with respect to elementary set E is denoted as $P(X|E)$. Since $\tilde{U}$ is the discussed set and events (assumptions or evidences) under consideration are only from it, we can consider $\tilde{U}$ instead of U approximately and then the prior probabilities or condition probabilities of target events can be estimated from $\tilde{U}$.

In the following, we will use the notation 'U' instead of '$\tilde{U}$' to express notions. Due to the sample set U is finite and nonempty in rough set theory, prior probability and condition probability of X are estimated [39,44] as follows:

$$P(X) = \frac{card(X)}{card(U)}, \quad P(X|E) = \frac{P(X \cap E)}{P(E)} = \frac{card(X \cap E)}{card(E)} \tag{1}$$

where $card(X)$ denotes the cardinality of set X. Condition probability $P(X|E)$ is also known as rough membership function. Since the occurrence of any target event is not certain, we have $0 < P(X) < 1$ and $0 < P(E) < 1$.

2.2. Decision table

If $A = C \cup D$ and $C \cap D = \emptyset$, C is a finite nonempty set of condition attributes, D is a finite nonempty set of decision attributes, then information system IS is called as a decision table, denoted as $DT = (U, C \cup D, V, \rho)$. With equivalence relation C, universe U can be partitioned into a collection of equivalence classes $U/C = \{C_1, C_2, \ldots, C_{card(U/C)}\}$. Similarly, universe U can be partitioned into another collection of equivalence classes $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$ with equivalence relation D. Each element of U/C is called a condition class and each element of U/D is called a decision class. Probability distributions defined on σ algebra of measurable subsets of universe U with respect to C and D are presented respectively as follows:

$$[U/C : P] = \begin{bmatrix} C_1 & C_2 & \cdots & C_{card(U/C)} \\ P(C_1) & P(C_2) & \cdots & P(C_{card(U/C)}) \end{bmatrix}$$

$$[U/D : P] = \begin{bmatrix} D_1 & D_2 & \cdots & D_{card(U/D)} \\ P(D_1) & P(D_2) & \cdots & P(D_{card(U/D)}) \end{bmatrix} \tag{2}$$

Obviously, $\sum_{i=1}^{card(U/C)} P(C_i) = 1$ and $\sum_{j=1}^{card(U/D)} P(D_j) = 1$. Given a decision table, decision attribute set D will not be changed in the sequent procedure of data analysis. So relevant probability distribution $[U/D : P]$ will not be changed, either. In addition, if there are multiple decision attributes, a single new decision attribute can be applied instead which values are presented as the combinations of multiple decision attribute values. Without loss of generality, we only consider $D = \{d\}$ in the discussions.

2.3. Certainty gain

Based on the notion of condition probability, the certainty gain function [46] is often used to evaluate the degree of increase or decrease of the prediction probability compared with the prior probability. For $X \subseteq U$, its certainty gain function with respect to elementary set E is defined as:

$$g(X|E) = \frac{P(X|E)}{P(X)} - 1 \tag{3}$$

If $g(X|E) > 0$, then $P(X|E) > P(X)$, it indicates that the probability of occurrence of X will increase according to event E. We can properly confirm that X will occur. If $g(X|E) < 0$, then $P(X|E) < P(X)$, it indicates that the probability of occurrence of X will decrease according to event E. In this case, we can properly confirm that X will not occur. If $g(X|E) = 0$, then $P(X|E) = P(X)$, it indicates that the probability of occurrence of X will not be affected by the added event E, so X and E can be considered to be independent, namely, evidence E has no effect on assumption X.

2.4. Absolute certainty gain

Under some circumstances, we are only interested in whether the probability of occurrence of event X will be affected by adding new evidence E regarding the value of prior probability $P(X)$ and how much the degree of the affection is. Absolute certainty gain [44,47] can be used to meet this requirement. For $X \subseteq U$, its absolute certainty gain function with respect to elementary set E is defined as:

$$gabs(X|E) = |P(X|E) - P(X)| \tag{4}$$

where $| * |$ denotes absolute value function. $gabs(X|E)$ only represents the absolute degree of change for the probability of occurrence of event X relative to its prior probability after adding new event E, irrespective of increase or decrease. $gabs(X|E) = 0$ means the probability of occurrence of X will not be affected by the added event E, namely, X and E are independent.

It is easy to get $gabs(X|E) = gabs(\neg X|E)$. It indicates that the new evidence E will have the same absolute affection both on X and its complement $\neg X$ relative to their prior probabilities.

3. The probabilistic evolution of rough set model

Some extended models have been proposed since original rough set model was introduced. In this section, we only concentrate on some probabilistic approaches to rough set model, especially on VPRSM and BRSM. Since these models can be directly derived from DTRSM, more details can be referenced in [26,27,29] from the perspective of decision theoretic framework.

3.1. Pawlak's rough set model

Typical objective of rough set theory is to form an approximate definition of the target set $X \subseteq U$ in terms of some definable sets, especially, when the target set is indefinable or vague. The upper and lower approximation of X with respect to equivalence relation A are denoted as $\overline{A}X$ and $\underline{A}X$, respectively, and defined as follows:

$$\overline{A}X = \cup\{E|E \cap X \neq \emptyset, E \in U/A\}, \quad \underline{A}X = \cup\{E|E \subseteq X, E \in U/A\} \tag{5}$$

The upper approximation of X is composed of objects that belong to set X possibly, and the lower approximation of X is composed of objects that belong to set X certainly. The upper and lower approximation of X approximate the concept X from two sides. In other words, the concept X can be described by two crisp (precise) sets approximately. Specially, if the concept X is uncertain or vague, such approximate descriptions have important meanings.

An equivalent formulation for upper and lower approximation with condition probability can be expressed as follows:

$$\begin{aligned} &\overline{A}X = \cup\{E|P(X|E) > 0, E \in U/A\} \\ &\underline{A}X = \cup\{E|P(X|E) = 1, E \in U/A\} \end{aligned} \tag{6}$$

Three kinds of approximation regions of X with respect to A can be defined according to its upper and lower approximation, respectively.

Positive region:

$$POS_A(X) = \underline{A}X = \cup\{E|P(X|E) = 1, E \in U/A\} \tag{7}$$

Negative region:

$$NEG_A(X) = U - \overline{A}X = \cup\{E|P(X|E) = 0, E \in U/A\} \tag{8}$$

Boundary region:

$$BND_A(X) = \overline{A}X - \underline{A}X = \cup\{E|0 < P(X|E) < 1, E \in U/A\} \tag{9}$$

Positive region depicts the set of objects that belong to X certainly, negative region depicts the set of objects that not belong to X certainly, and boundary region is composed of objects that we can not discern whether they are included or not included in X.

Proposition 1. *Suppose an information system* $IS = (U, A, V, \rho)$, $X \subseteq U$, $\forall B \subset A$, *if* $E_i, E_j \in U/A (i \neq j)$, $F \in U/B$ *and* $F = E_i \cup E_j$, *we have:*

(1) *if* $E_i \subseteq POS_A(X)$ *and* $E_j \subseteq POS_A(X)$, *then* $F \subseteq POS_B(X)$;
(2) *if* $E_i \subseteq NEG_A(X)$ *and* $E_j \subseteq NEG_A(X)$, *then* $F \subseteq NEG_B(X)$;
(3) *if* $E_i \subseteq BND_A(X)$ *and* $E_j \subseteq BND_A(X)$, *then* $F \subseteq BND_B(X)$;
(4) *if* $E_i \subseteq POS_A(X)$ *and* $E_j \subseteq BND_A(X)$, *then* $F \subseteq BND_B(X)$;
(5) *if* $E_i \subseteq NEG_A(X)$ *and* $E_j \subseteq BND_A(X)$, *then* $F \subseteq BND_B(X)$;
(6) *if* $E_i \subseteq POS_A(X)$ *and* $E_j \subseteq NEG_A(X)$, *then* $F \subseteq BND_B(X)$.

Proof. Only the proofs for case (6) are given, the others can be proved similarly.

Because $E_i \subseteq POS_A(X)$, $E_j \subseteq NEG_A(X)$ and E_i and E_j are elementary sets with respect to A, so $E_i \subseteq X$ and $E_j \cap X = \emptyset$, thus $(E_i \cup E_j) \cap X \neq \emptyset$ and $(E_i \cup E_j) \not\subset X$, it has $F \subseteq BND_B(X)$. □

After removing some attributes from the entire set of description attributes, some elementary sets will be merged, and the approximation regions of X will be changed under the new set of attributes. Proposition 1 presents all merging cases between two elementary sets during the procedures of attribute reduction.

Proposition 2. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$, $\forall D_p, D_q \in U/D (p \neq q)$, then $POS_C(D_p) \cap POS_C(D_q) = \emptyset$.*

Proof. $\forall C_i \in U/C$, if $C_i \subseteq POS_C(D_p)$, then $C_i \subseteq D_p$. Since $D_p, D_q \in U/D$, thus $D_p \cap D_q = \emptyset$, So $C_i \cap D_q = \emptyset$, it has $C_i \cap POS_C(D_q) = \emptyset$. Similarly, $\forall C_i \in U/C$, if $C_i \subseteq POS_C(D_q)$, then $C_i \cap POS_C(D_p) = \emptyset$. Consequently, $POS_C(D_p) \cap POS_C(D_q) = \emptyset$. □

Proposition 2 indicates that any condition class will be included in the positive region of at most one decision class.

3.2. Variable precision rough set model

In practical applications, Pawlak's rough set model can not deal with data sets which have some noisy data effectively. Lots of information in the boundary region will be abandoned which may provide latent useful knowledge. By applying the parameter, the approximate regions can be adjusted and controlled in VPRSM. Given a parameter β, $0 \leq 1 - \beta < P(X) < \beta \leq 1$, three kinds of β-approximation regions of concept $X \subseteq U$ with respect to equivalence relation A can be defined as follows:

β-positive region:

$$POS_A^{\beta}(X) = \cup\{E | P(X|E) \geq \beta, E \in U/A\} \tag{10}$$

β-negative region:

$$NEG_A^{\beta}(X) = \cup\{E | P(X|E) \leq 1 - \beta, E \in U/A\} \tag{11}$$

β-boundary region:

$$BND_A^{\beta}(X) = \cup\{E | 1 - \beta < P(X|E) < \beta, E \in U/A\} \tag{12}$$

β-positive region is the collection of all those elementary sets which can be included in X with the certainty degree not lower than β. β-negative region is composed of all those elementary sets which can be included in the complement of X, viz. $\neg X$, with the certainty degree not lower than β. β-boundary region is composed of all those elementary sets which can not be classified into X and its complement $\neg X$ with the certainty degree not lower than β.

Due to $1 - \beta < \beta$, so it implies $0.5 < \beta \leq 1$. It indicates that $0 \leq 1 - \beta < P(X) < \beta \leq 1$ includes the original 0.5 symmetric variable precision rough set model [17]. When $\beta = 1$, the VPRSM will degenerate to the Pawlak's rough set model.

Remarks. Absolute certainty gain function is used to describe the β-approximation regions of $X \subseteq U$ with $\beta > P(X)$ in [44,47] as follows:

$$POS_A^{\beta}(X) = \cup\{E : gabs(X|E) \geq \beta - P(X)\} \tag{13}$$

$$NEG_A^{\beta}(X) = \cup\{E : gabs(\neg X|E) \geq \beta - P(X)\} \tag{14}$$

and then a unified description for the β-positive region of X and its complement $\neg X$ is also given as:

$$POS_A^{\beta}(X, \neg X) = \cup\{E : gabs(X|E) \geq \beta - P(X)\} \tag{15}$$

After being analyzed critically, it is found that these notions are unreasonable. A counterexample can be given as follows:

Suppose $P(X|E) = \frac{1}{2}$, $P(X) = \frac{2}{3}$ and $\beta = \frac{3}{4}$. Obviously, $\beta > P(X)$ and we have $|P(X|E) - P(X)| = |P(\neg X|E) - P(\neg X)| > \beta - P(X)$. According to formulas (13) and (14), $E \subseteq POS_A^{\beta}(X)$ and $E \subseteq NEG_A^{\beta}(X)$, it is a contradiction. Furthermore, according to formula (15), we have $E \subseteq POS_A^{\beta}(X, \neg X)$. However, $1 - \beta < P(X|E) < \beta$, according to the definition of β-boundary region, E should be included in $BND_A^{\beta}(X)$. Similarly, E should also be included in $BND_A^{\beta}(\neg X)$. So the descriptions of β-approximation regions of $X \subseteq U$ in terms of absolute certainty gain function are unreasonable.

Proposition 3. *Suppose an information system* $IS = (U, A, V, \rho), X \subseteq U$, *given* $0 \leq 1 - \beta < P(X) < \beta \leq 1, \forall E \in U/A$, *we have:*

(1) *if* $E \subseteq POS_A^\beta(X)$, *then* $gabs(X|E) \geq \beta - P(X)$;
(2) *if* $E \subseteq NEG_A^\beta(X)$, *then* $gabs(\neg X|E) \geq \beta - P(\neg X)$.

Proof. (1) because $E \subseteq POS_A^\beta(X)$ and $\beta > P(X)$, so $P(X|E) \geq \beta > P(X)$. It has $P(X|E) - P(X) \geq \beta - P(X) > 0$, namely, $gabs(X|E) \geq \beta - P(X)$.
(2) can be proved similarly. □

Proposition 3 means that formula (13) is just an implication, not an equality.

3.3. Bayesian rough set model

The value of parameter β in VPRSM is often difficult to choose in real applications and the optimal β value can not be given objectively. The extracted decision rule sets will be distinct based on different β values. Slezak and Ziarko put forward BRSM [23] in which the prior probability of the event under consideration is chosen as a benchmark value. BRSM is a hybrid product which connects rough set theory and Bayesian reasoning validly and reasonably. It is more appropriate to application problems concerned with achieving any certainty gain during the procedures of prediction or decision making rather than meeting a special certainty goal [34].

In BRSM, three kinds of B-approximation regions of concept $X \subseteq U$ with respect to equivalence relation A can be defined as follows:

B-positive region:

$$POS_A^*(X) = \cup\{E|P(X|E) > P(X), E \in U/A\} \tag{16}$$

B-negative region:

$$NEG_A^*(X) = \cup\{E|P(X|E) < P(X), E \in U/A\} \tag{17}$$

B-boundary region:

$$BND_A^*(X) = \cup\{E|P(X|E) = P(X), E \in U/A\} \tag{18}$$

B-positive region can be interpreted as the collection of all those elementary sets of U/A which can increase the probability of occurrence of X relative to its prior probability $P(X)$. In other words, the certainty degree of occurrence of X will increase after adding any new event in its B-positive region, and then we would rather believe that event X will occur.

B-negative region can be interpreted as the collection of all those elementary sets of U/A which can decrease the probability of occurrence of X relative to its prior probability $P(X)$. In other words, the certainty degree of occurrence of X will decrease after adding any new event in its B-negative region, and then we would rather believe that event X will not occur.

Any elementary sets of U/A in the B-boundary region are independent with X, namely, newly added information has no effect on the prediction probability of X relative to its prior probability. When assumption X is under consideration, the evidence in its B-boundary region will not be useful.

In practice, positive evidences are beneficial for decision making. Taking an example of medical diagnosis, the types of diseases can not be confirmed only by the symptoms, sometimes. However, we can give a finally certain diagnosis after getting the test results from medical equipments. On the contrary, negative evidences are also beneficial for applications. There may be some potential fault points in the processes of fault detection. After making some professional tests, we can eliminate impossible points as much as possible and finally, a certain result can be reported.

Proposition 4. *Suppose an information system* $IS = (U, A, V, \rho), X \subseteq U, \forall B \subset A$, *if* $E_i, E_j \in U/A (i \neq j), F \in U/B$ *and* $F = E_i \cup E_j$, *then we have:*

(1) *if* $E_i \subseteq POS_A^*(X)$ *and* $E_j \subseteq POS_A^*(X)$, *then* $F \subseteq POS_B^*(X)$;
(2) *if* $E_i \subseteq NEG_A^*(X)$ *and* $E_j \subseteq NEG_A^*(X)$, *then* $F \subseteq NEG_B^*(X)$;
(3) *if* $E_i \subseteq BND_A^*(X)$ *and* $E_j \subseteq BND_A^*(X)$, *then* $F \subseteq BND_B^*(X)$;
(4) *if* $E_i \subseteq POS_A^*(X)$ *and* $E_j \subseteq BND_A^*(X)$, *then* $F \subseteq POS_B^*(X)$;
(5) *if* $E_i \subseteq NEG_A^*(X)$ *and* $E_j \subseteq BND_A^*(X)$, *then* $F \subseteq NEG_B^*(X)$.

Proof. Only the proofs for case (1) are given, the others can be proved similarly.

Because $E_i \subseteq POS_A^*(X)$ and $E_j \subseteq POS_A^*(X)$, so $P(X|E_i) > P(X)$ and $P(X|E_j) > P(X)$, then $P(X \cap E_i) > P(X)P(E_i)$ and $P(X \cap E_j) > P(X)P(E_j)$.

For F, due to $E_i \cap E_j = \emptyset$, it has:

$P(X|F) = \frac{P(X\cap F)}{P(F)} = \frac{P(X\cap(E_i\cup E_j))}{P(E_i\cup E_j)} = \frac{P(X\cap E_i)+P(X\cap E_j)}{P(E_i)+P(E_j)} > P(X)$.
Consequently, $F \subseteq POS^*_B(X)$. □

Compared with Pawlak's rough set model, case (6) in the Proposition 1 is not satisfied in the BRSM. If $E_i \subseteq POS^*_A(X)$ and $E_j \subseteq NEG^*_A(X)$, we can not determine which B-approximation region will include $E_i \cup E_j$ after reducing some attributes. All of the three kinds of B-approximation regions are possible. In addition, Proposition 2 is also not satisfied in BRSM. An elementary set can be included in the B-positive region of more than one decision class.

Given a decision table $DT = (U, C \cup D, V, \rho)$, the B-positive region of decision attribute set D with respect to condition attribute set C is defined as follows:

$$POS^*_C(DT) = \cup_{D_j\in U/D} POS^*_C(D_j) \tag{19}$$

Any elementary sets in $POS^*_C(DT)$ will increase the probability of occurrence of at least one decision class relative to its prior probability.

For any $C_i \in U/C$, if $\forall D_j \in U/D$, it has $C_i \subseteq POS^*_C(D_j)$, then C_i is considered to be included in the absolute positive region of decision table DT, denoted as $APOS^*_C(DT)$; if $\forall D_j \in U/D$, it has $C_i \subseteq NEG^*_C(D_j)$, then C_i is considered to be included in the absolute negative region of decision table DT, denoted as $ANEG^*_C(DT)$ and if $\forall D_j \in U/D$, it has $C_i \subseteq BND^*_C(D_j)$, then C_i is considered to be included in the absolute boundary region of decision table DT, denoted as $ABND^*_C(DT)$. Formally, they can be described as follows:

$$APOS^*_C(DT) = \cup\{C_i|C_i \in U/C, \forall D_j \in U/D \rightarrow C_i \subseteq POS^*_C(D_j)\} \tag{20}$$

$$ANEG^*_C(DT) = \cup\{C_i|C_i \in U/C, \forall D_j \in U/D \rightarrow C_i \subseteq NEG^*_C(D_j)\} \tag{21}$$

$$ABND^*_C(DT) = \cup\{C_i|C_i \in U/C, \forall D_j \in U/D \rightarrow C_i \subseteq BND^*_C(D_j)\} \tag{22}$$

Proposition 5. *$APOS^*_C(DT) = \emptyset$, $ANEG^*_C(DT) = \emptyset$.*

Proof. Suppose $\exists C_i \in U/C$, such that $C_i \subseteq APOS^*_C(DT)$. Then $\forall D_j \in U/D$, it has $P(D_j|C_i) > P(D_j)$, so $\sum_{j=1}^{card(U/D)} P(D_j|C_i) > \sum_{j=1}^{card(U/D)} P(D_j)$. With $\sum_{j=1}^{card(U/D)} P(D_j|C_i) = 1$ and $\sum_{j=1}^{card(U/D)} P(D_j) = 1$, it is a contradiction.

Suppose $\exists C_i \in U/C$, such that $C_i \subseteq ANEG^*_C(DT)$. Then $\forall D_j \in U/D$, it has $P(D_j|C_i) < P(D_j)$, so $\sum_{j=1}^{card(U/D)} P(D_j|C_i) < \sum_{j=1}^{card(U/D)} P(D_j)$. With $\sum_{j=1}^{card(U/D)} P(D_j|C_i) = 1$ and $\sum_{j=1}^{card(U/D)} P(D_j) = 1$, it is also a contradiction. □

Proposition 5 indicates that none of the elementary sets will increase or decrease the probabilities of occurrence of all decision classes relative to their prior probabilities. However, the absolute boundary region may not be empty. We are not interested in the objects belonging to the absolute boundary region of a decision table because the probability of occurrence of each decision class will not be affected by the newly added information in this approximation region.

Proposition 6. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$, if $ABND^*_C(DT) \neq U$, then $\exists C_i \in U/C$ and $\exists D_p, D_q \in U/D (p \neq q)$, it has $C_i \subseteq POS^*_C(D_p)$ and $C_i \subseteq NEG^*_C(D_q)$.*

Proof. Since $ABND^*_C(DT) \neq U$, then $\exists C_i \in U/C$, such that $C_i \subseteq U - ABND^*_C(DT)$. So $\exists D_p \in U/D$, it has $C_i \not\subset BND^*_C(D_p)$.

Suppose $C_i \subseteq POS^*_C(D_p)$ and $\forall D_j \in U/D$ ($j \neq p$), such that $C_i \not\subset NEG^*_C(D_j)$. It has $P(D_p|C_i) > P(D_p)$, so $\sum_{j=1}^{card(U/D)} P(D_j|C_i) = \sum_{j=1\wedge j\neq p}^{card(U/D)} P(D_j|C_i) + P(D_p|C_i) > \sum_{j=1}^{card(U/D)} P(D_j)$. With $\sum_{j=1}^{card(U/D)} P(D_j|C_i) = 1$ and $\sum_{j=1}^{card(U/D)} P(D_j) = 1$, it is a contradiction. Consequently, $\exists D_q \in U/D(q \neq p)$, such that $C_i \subseteq NEG^*_C(D_q)$.

Similarly, suppose $C_i \subseteq NEG^*_C(D_p)$, it has $\exists D_q \in U/D(q \neq p)$, such that $C_i \subseteq POS^*_C(D_q)$. □

If there is an elementary set in U/C that is not included in the absolute boundary region, then it will increase the probabilities of occurrence of some decision classes relative to their prior probabilities and meanwhile, decrease the probabilities of occurrence of some other decision classes relative to their probabilities. In this case, we can rather believe that some assumptions will occur as well as some other assumptions will not occur. Proposition 6 implies that if not all elementary sets under equivalence relation C are included in the absolute boundary region, then the classification knowledge under C will be beneficial for decision making.

4. Attribute reduction in BRSM for binary decision problems

For binary decision problems in BRSM, without loss of generality, we denote $U/D = \{X, \neg X\}$ for decision tables which have two decision classes. The B-approximation regions of X and its complement $\neg X$ have properties as follows [34,43]:

Property 1 (Symmetry(Duality))**.**

$$POS_C^*(X) = NEG_C^*(\neg X) \quad and \quad NEG_C^*(X) = POS_C^*(\neg X) \tag{23}$$

Property 2 (Equivalence)**.**

$$ABND_C^*(X) = BND_C^*(X) = BND_C^*(\neg X) \tag{24}$$

According to the Properties 1 and 2, if the B-approximation regions of X have been obtained, then the information about the B-approximation regions of its complement $\neg X$ can be captured simultaneously.

Slezak [43] have utilized the global relative gain function to construct attribute reduction model in BRSM for binary decision problems. Based on certainty gain function, the local relative gain function is defined as:

$$r(X|E) = max\{g(X|E), g(\neg X|E)\} \tag{25}$$

and global relative gain function is defined as:

$$R(X|A) = \sum_{E \in U/A} P(E)r(X|E) \tag{26}$$

The local relative gain function reflects the relative certainty increase of the occurrence of X or $\neg X$ after adding new information. Due to the duality of regions with respect to the events and their complements, if the certainty of occurrence of one decision class increases, then the certainty of occurrence of the other decision class will decrease. In other cases, newly added information will have no effect both on X and $\neg X$. The global relative gain function reflects the average relative certainty gain over all elementary sets under attribute set A.

Theorem 1 [43]**.** *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with binary decision classes, $U/D = \{X, \neg X\}$, $\forall B \subseteq C$, it has $R(X|B) \le R(X|C)$ and equality holds, if and only if $POS_C^*(X) \subseteq POS_B^*(X)$ and $POS_C^*(\neg X) \subseteq POS_B^*(\neg X)$.*

According to Property 1 (see formula (23)), equality holds in Theorem 1, if and only if $POS_C^*(X) \subseteq POS_B^*(X)$ and $NEG_C^*(X) \subseteq NEG_B^*(X)$.

After some attributes being removed from the entire condition attribute set, some former elementary sets may be merged. According to Proposition 4, the current B-positive region of X will not be reduced by merging the former elementary sets which are included in the former B-boundary region and the former B-positive region, respectively. Similarly, the current B-negative region will not be reduced by merging the former elementary sets which are included in the former B-boundary region and the former B-negative region, respectively. Essentially, the objects belonging to $POS_C^*(X)$ and $NEG_C^*(X)$ respectively will be discerned by Theorem 1.

Ziarko [44] have also applied the normalized expected gain function to construct attribute reduction model in BRSM for binary decision problems. Expected gain function is defined as:

$$egabs(X|A) = \sum_{E \in U/A} P(E)gabs(X|E) \tag{27}$$

Since $gabs(X|E) = gabs(\neg X|E)$ for any elementary set $E \in U/A$, it has $egabs(X|A) = egabs(\neg X|A)$. The expected gain function measures the average absolute degree of increase or decrease of the occurrence probability of X or $\neg X$ under the new classification knowledge.

Since $0 \le egabs(X|A) \le 2P(X)(1 - P(X))$[44], the normalized expected gain function is defined as: $\lambda(X|A) = \frac{egabs(X|A)}{2P(X)(1-P(X))}$, then $\lambda(X|A) \in [0, 1]$. $\lambda(X|A) = 1$ only if X is definable under the classification knowledge U/A.

Theorem 2. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with binary decision classes, $U/D = \{X, \neg X\}$, $\forall B \subseteq C$, it has $\lambda(X|B) \le \lambda(X|C)$ and equality holds, if and only if $POS_C^*(X) \subseteq POS_B^*(X)$ and $NEG_C^*(X) \subseteq NEG_B^*(X)$.*

According to Theorems 1 and 2, both of the global relative gain function and expected gain function discern the objects between the B-positive region and B-negative region of X (or $\neg X$). So attribute reduction model for binary decision problems in BRSM based on the global relative gain function or normalized expected gain function are equivalent and an unified model can be described as follows.

Let $\varphi_2 \in \{R, \lambda\}$, suppose a decision table $DT = (U, C \cup D, V, \rho)$ with binary decision classes, $U/D = \{X, \neg X\}$. $\forall B \subseteq C$, if B satisfies the following two criteria:

(1) $\varphi_2(X|B) = \varphi_2(X|C)$;
(2) for $\forall b \in B$, it has $\varphi_2(X|B - \{b\}) < \varphi_2(X|B)$.

Then B is called a reduct of C with respected to D, or called a reduct briefly.

Condition (1) indicates the joint sufficiency of the attribute set B, namely, attribute set B is sufficient to preserve the property R or λ. Condition (2) means each element in B is individually necessary as remaining the property. Discernibility matrices and discernibility functions [45] are often used to describe the knowledge in rough set theory and then the set of reducts can be obtained based on Boolean reasoning. In BRSM, discernibility matrices can also be constructed for binary decision problems.

Given a decision table $DT = (U, C \cup D, V, \rho)$ with binary decision classes, $U/D = \{X, \neg X\}$, $\forall x \in U$, its characteristic function with respect to X can be defined as:

$$\delta_X(x) = \begin{cases} 1 & x \in POS_C^*(X) \\ 0 & x \in BND_C^*(X) \\ -1 & x \in NEG_C^*(X) \end{cases} \tag{28}$$

For a decision table with two decision classes, its discernibility matrix can be defined as a $n \times n$ matrix $DM_2(DT) = (c_{ij})_{n\times n}$, where the element c_{ij} satisfies:

$$c_{ij} = \begin{cases} \{a | a \in C \wedge \rho(x_i, a) \neq \rho(x_j, a)\} & 1 \leq j < i \leq n,\ \delta_X(x_i)\delta_X(x_j) = -1 \\ \emptyset & others \end{cases} \tag{29}$$

According to the properties of symmetry and equivalence for the B-approximation regions of X and $\neg X$, the discernibility matrix $DM_2(DT) = (c_{ij})_{n\times n}$ will be the same irrespective of the characteristic functions of objects with respect to X or its complement $\neg X$.

Theorem 3. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with binary decision classes, $U/D = \{X, \neg X\}$, $\forall B \subseteq C$, B is a reduct if and only if B satisfies the following two criteria:*

(1) *$\forall i, j$, $(1 \leq j < i \leq n)$, if $c_{ij} \neq \emptyset$, then $B \cap c_{ij} \neq \emptyset$;*
(2) *$\forall b \in B$, $\exists i, j$, $c_{ij} \neq \emptyset$, such that $(B - \{b\}) \cap c_{ij} = \emptyset$.*

Theorem 3 indicates that the two kinds of current attribute reduction models in BRSM for binary decision problems can be constructed by an unified version based on the proposed discernibility matrix, no matter whether the attribute reduct is defined in terms of the global relative gain function or the normalized expected gain function.

5. Approximation quality measures for multi-decision problems

In practical applications, we often face decision problems which have multi-decision classes. Nishino et al applied the VPBRSM to extract multi-decision rules based on a newly proposed information gain [36]. Slezak discussed the RBM's capability for dealing with multi-decision problems in which a parameter matrix needs to be predefined [39]. Yao et al also studied the DTRSM for handling multi-decision problems according to the decision theoretic framework [30,31]. BRSM for multi-decision problems has been investigated by Ziarko firstly, and the proposed multi-valued expected gain function is used to construct an attribute reduction model [48].

5.1. Multi-valued expected gain function

Given a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$, the multi-valued expected gain function is defined as:

$$megabs(D|C) = \sum_{D_j \in U/D} P(D_j) egabs(D_j|C) \tag{30}$$

$megabs(D|C)$ reflects the average measure for the absolute change of occurrence probabilities over all decision classes under classification knowledge U/C relative to their prior probabilities.

Proposition 7. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$, $U/C = \{C_1, C_2, \ldots, C_{card(U/C)}\}$, it has:*

$$megabs(D|C) = \sum_{D_j \in U/D} P(D_j) \sum_{C_i \in U/C} \left| P(D_j \cap C_i) - P(D_j)P(C_i) \right|.$$

Proof. It follows the definition of expected gain function directly. □

Multi-valued expected gain function can also be given another description for computation.

Proposition 8. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$, $U/C = \{C_1, C_2, \ldots, C_{card(U/C)}\}$, it has:*

$$megabs(D|C) = \sum_{D_j \in U/D} P^2(D_j) \sum_{C_i \in U/C} gabs(C_i|D_j).$$

Proof. $\forall C_i \in U/C$, and $\forall D_j \in U/D$, it has $P(C_i)\left|P(D_j|C_i) - P(D_j)\right| = \left|P(D_j \cap C_i) - P(C_i)P(D_j)\right| = P(D_j)\left|P(C_i|D_j) - P(C_i)\right|$.
So $\sum_{D_j \in U/D} P(D_j) \sum_{C_i \in U/C} P(C_i)\left|P(D_j|C_i) - P(D_j)\right| = \sum_{D_j \in U/D} P(D_j) \sum_{C_i \in U/C} P(D_j)\left|P(C_i|D_j) - P(C_i)\right|$.
Consequently, $megabs(D|C) = \sum_{D_j \in U/D} P^2(D_j) \sum_{C_i \in U/C} gabs(C_i|D_j)$. □

Proposition 9 [48]. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, it has* $0 \le megabs(D|C) \le 2\sum_{D_j \in U/D} P^2(D_j)(1 - P(D_j))$.

Since the decision attribute set D will not be changed during the procedures of data analysis, the value of $megabs(D|C)$ can be normalized into the interval [0, 1] by dividing the constant $2\sum_{D_j \in U/D} P^2(D_j)(1 - P(D_j))$ according to the Proposition 9.

5.2. Multi-valued expected total gain function

The probabilities of occurrence of all decision classes may be affected by adding a new evidence. Sometimes, we are interested in the average measure for the total affection caused by each elementary set belonging to U/C.

For each $C_i \in U/C$, its total affection on all decision classes can be defined as :

$$tgabs(D|C_i) = \sum_{D_j \in U/D} gabs(D_j|C_i) \tag{31}$$

So the multi-valued expected total gain function can be defined as:

$$metgabs(D|C) = \sum_{C_i \in U/C} P(C_i) \sum_{D_j \in U/D} gabs(D_j|C_i) \tag{32}$$

Given a decision table, the estimation of prior probability of each decision class will not be changed under different sets of condition attributes. Multi-valued expected total gain function can be considered as a simplification of multi-valued expected gain function to some extent, but they have different meanings. For each elementary set belonging to U/C, the multi-valued expected total gain function is focused on its total affection on all decision classes and the multi-valued expected gain function is focused on the average affection over all decision classes.

Proposition 10. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D}\}$, $U/C = \{C_1, C_2, \ldots, C_{card(U/C)}\}$, it has:*

$$metgabs(D|C) = \sum_{D_j \in U/D} \sum_{C_i \in U/C} \left|P(D_j \cap C_i) - P(D_j)P(C_i)\right|.$$

Proof. it follows the definition of absolute gain function directly. □

According to Proposition 7, $megabs(D|C)$ is considered as the measure of average deviation from probabilistic independence between each elementary set and each decision class. Therefore, according to Proposition 10, $metgabs(D|C)$ can be considered as the measure of total deviation from probabilistic independence between each elementary set and each decision class.

Proposition 11. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$, $U/C = \{C_1, C_2, \ldots, C_{card(U/C)}\}$, it has:*

$$metgabs(D|C) = metgabs(C|D) = \sum_{D_j \in U/D} P(D_j) \sum_{C_i \in U/C} gabs(C_i|D_j).$$

Proof. Because $P(C_i)gabs(D_j|C_i) = P(D_j)gabs(C_i|D_j)$ for any $C_i \in U/C$ and any $D_j \in U/D$. So $metgabs(D|C) = \sum_{C_i \in U/C} \sum_{D_j \in U/D} P(C_i)gabs(D_j|C_i) = \sum_{D_j \in U/D} P(D_j) \sum_{C_i \in U/C} gabs(C_i|D_j)$, namely, $metgabs(D|C) = metgabs(C|D)$. □

Proposition 11 indicates that classification knowledge U/D and U/C have duality properties based on the multi-valued expected total gain function to some extent. In this case, the affection on decision classification U/D caused by the classification knowledge U/C is equal to the affection on condition classification U/C caused by the classification knowledge U/D. This can also be found from the Proposition 10, the deviation from probabilistic independence between each condition class and each decision class is mutual and symmetric.

Remarks. There may be some clerical errors in paper [48]. Propositions 10 and 11 should be satisfied under the definition of multi-valued expected total gain function, not the definition of multi-valued expected gain function as presented in paper [48].

Proposition 12. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, it has $0 \leq metgabs(D|C) \leq 2\left(1 - \frac{1}{n}\right)$, where n is the number of objects in U.*

Proof. obviously, $metgabs(D|C) \geq 0$ and equality holds, if and only if $P(C_i \cap D_j) = P(C_i)P(D_j)$ holds for each elementary set and each decision class, namely, each elementary set and each decision class are independent.

When $\forall D_j \in U/D$ is definable with respect to the classification knowledge U/C, $metgabs(D|C)$ will has the maximal value. So it has:

$$\begin{aligned} metgabs(D|C) &= \sum_{D_j \in U/D} \sum_{C_i \in U/C} P(C_i) \left| P(D_j|C_i) - P(D_j) \right| \leq \sum_{D_j \in U/D} \left(\sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq D_j}} P(C_i)(1 - P(D_j)) + \sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq \neg D_j}} P(C_i)P(D_j) \right) \\ &= \sum_{D_j \in U/D} \left((1 - P(D_j))P(D_j) + P(D_j)P(\neg D_j)\right) = 2 \sum_{D_j \in U/D} (1 - P(D_j))P(D_j) = 2\left(1 - \sum_{D_j \in U/D} P^2(D_j)\right). \end{aligned}$$

When each decision class includes only one object in U, namely, all the objects in U have different decisions, $\sum_{D_j \in U/D} P^2(D_j)$ will have the minimal value $\frac{1}{n}$, so $metgabs(D|C) \leq 2\left(1 - \frac{1}{n}\right)$. □

According to the Proposition 12, the value of $metgabs(D|C)$ can be normalized into the interval [0, 1] by dividing the constant $2\left(1 - \frac{1}{n}\right)$. Compared with the Proposition 9, it need not to calculate the prior probability of each decision class as normalizing the multi-valued expected total gain function. Moreover, it is easy to obtain the value of n from the given decision table.

5.3. Positive and negative multi-valued expected gain function

Absolute gain function can not reflect the trend of change of the occurrence probabilities of assumptions relative to their prior probabilities after adding new information. Increase or decrease, it can not be measured. As discussed in Section 5.3, both the increase (positive) and decrease (negative) of the occurrence probabilities of target events will be beneficial for decision making, so the trend of change of the occurrence probabilities of assumptions should be considered and measured.

Positive multi-valued expected gain function can be defined as:

$$Pmeg(D|C) = \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq POS_C^*(D_j)}} (P(D_j|C_i) - P(D_j)) \tag{33}$$

If $ABND_C^*(DT) = U$, then set $Pmeg(D|C) = 0$. In this case, the newly added classification information is not useful for decision making relative to our prior knowledge.

Similarly, negative multi-valued expected gain function can be defined as:

$$Nmeg(D|C) = \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq NEG_C^*(D_j)}} (P(D_j) - P(D_j|C_i)) \tag{34}$$

If $ABND_C^*(DT) = U$, then also set $Nmeg(D|C) = 0$. According to the Proposition 6, if $ABND_C^*(DT) \neq U$, $POS_C^*(DT)$ is not equal to empty set. In this case, the values of $Pmeg(D|C)$ and $Nmeg(D|C)$ must be greater than zero.

Positive multi-valued expected gain function is only focused on the degree of increase of the occurrence probabilities over all decision classes after adding new information relative to their prior probabilities. Negative multi-valued expected gain function is only focused on the degree of decrease of the occurrence probabilities over all decision classes after adding new information relative to their prior probabilities.

Proposition 13. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, it has* $0 \le Pmeg(D|C) \le 1 - \frac{1}{n}$ *and* $0 \le Nmeg(D|C) \le 1 - \frac{1}{n}$*, where n is the number of objects in U.*

Proof. Only the first one is proved. The second one can be proved similarly.

Obviously, $Pmeg(D|C) \ge 0$. When $\forall D_j \in U/D$ is definable with respect to the classification knowledge U/C, $Pmeg(D|C)$ will has the maximal value. So we can have:

$$\begin{aligned} Pmeg(D|C) &\le \sum_{D_j \in U/D} \sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq D_j}} P(C_i)(1 - P(D_j)) = \sum_{D_j \in U/D} (1 - P(D_j)) \sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq D_j}} P(C_i) \\ &= \sum_{D_j \in U/D} (1 - P(D_j))P(D_j) = 1 - \sum_{D_j \in U/D} P^2(D_j) \le 1 - \frac{1}{n}. \quad \square \end{aligned}$$

According to the Proposition 13, the values of $Pmeg(D|C)$ and $Nmeg(D|C)$ can be normalized into the interval [0, 1] by dividing the constant $1 - \frac{1}{n}$.

Proposition 14. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, it has $metgabs(D|C) = Pmeg(D|C) + Nmeg(D|C)$.*

Proof. $\forall C_i \in U/C$, it has:

$$P(C_i) \sum_{D_j \in U/D} gabs(D_j|C_i) = P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq POS_C^*(D_j)}} \big(P(D_j|C_i) - P(D_j)\big) + P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq NEG_C^*(D_j)}} \big(P(D_j) - P(D_j|C_i)\big).$$

So, $\sum_{C_i \in U/C} P(C_i) \sum_{D_j \in U/D} gabs(D_j|C_i) = \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq POS_C^*(D_j)}} (P(D_j|C_i) - P(D_j))$ $+ \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq NEG_C^*(D_j)}} (P(D_j) - P(D_j|C_i))$.

Namely, $metgabs(D|C) = Pmeg(D|C) + Nmeg(D|C)$. $\square$

Proposition 14 indicates that multi-valued expected total gain function can be divided into two parts, one is "positive" for the prediction probabilities of decision classes, and the other is "negative" for the prediction probabilities of decision classes according to the new classification knowledge.

5.4. Positive and negative multi-valued expected degree of certainty

$\forall C_i \in U/C$, if $C_i \subseteq POS_C^*(D_j)$, $P(D_j|C_i) - P(D_j)$ only measures the degree of increase of the occurrence probability of D_j relative to its prior probability under new event C_i. However, it can not measure the degree of increase of the certainty that D_j will occur. If the prior probability $P(D_j)$is higher, then $P(D_j|C_i) - P(D_j)$ will be relative small even $P(D_j|C_i) = 1$. On the other side, if $P(D_j)$ is lower, then $P(D_j|C_i) - P(D_j)$ may be relative high even $P(D_j|C_i) < 1$. In the reality, $P(D_j|C_i) = 1$ is more significant for us in despite of $P(D_j|C_i) - P(D_j)$ is small, because we can make a certain decision under this situation.

Since the trivial subsets of universe U are not interested in, so $P(D_j) = 1$ and $P(D_j) = 0$ will not be considered. Positive multi-valued expected degree of certainty can be defined as:

$$DPmeg(D|C) = \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq POS_C^*(D_j)}} \left(\frac{P(D_j|C_i) - P(D_j)}{1 - P(D_j)}\right) \tag{35}$$

If $ABND_C^*(DT) = U$, then set $DPmeg(D|C) = 0$.

If $P(D_j|C_i) > P(D_j)$, $\frac{P(D_j|C_i) - P(D_j)}{1 - P(D_j)}$ reflects the incremental degree of belief. It measures the degree of increase of the certainty that assumption D_j will be true under evidence C_i. In other words, it indicates the degree of decrease of the certainty that assumption D_j will be false. $1 - P(D_j)$ can be considered as the total degree that D_j will be false.

Similarly, negative multi-valued expected degree of certainty can be defined as:

$$DNmeg(D|C) = \sum_{\substack{C_i \in U/C \\ \wedge C_i \not\subseteq ABND_C^*(DT)}} P(C_i) \sum_{\substack{D_j \in U/D \\ \wedge C_i \subseteq NEG_C^*(D_j)}} \left(\frac{P(D_j) - P(D_j|C_i)}{P(D_j)}\right) \tag{36}$$

If $ABND_C^*(DT) = U$, then also set $DNmeg(D|C) = 0$.

If $P(D_j|C_i) < P(D_j)$, $\frac{P(D_j)-P(D_j|C_i)}{P(D_j)}$ reflects the incremental degree of unbelief. It measures the degree of increase of the certainty that assumption D_j will be false under evidence C_i. In other words, it indicates the degree of decrease of the certainty that assumption D_j will be true. $P(D_j)$ can be considered as the total degree that D_j will be true.

Proposition 15. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, it has $0 \leq DPmeg(D|C) \leq 1$ and $0 \leq DNmeg(D|C) \leq 1$.*

Proof. only the first one is proved. The second one can be proved similarly.

Obviously, $DPmeg(D|C) \geq 0$. When $\forall D_j \in U/D$ is definable with respect to the classification knowledge U/C, $DPmeg(D|C)$ will has the maximal value. So it has:

$$DPmeg(D|C) \leq \sum_{D_j \in U/D} \sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq D_j}} P(C_i)\left(\frac{1-P(D_j)}{1-P(D_j)}\right) = \sum_{D_j \in U/D} \sum_{\substack{C_i \in U/C \\ \wedge C_i \subseteq D_j}} P(C_i) = \sum_{D_j \in U/D} P(D_j) = 1. \quad \square$$

The values of $DPmeg(D|C)$ and $DNmeg(D|C)$ do not need to be normalized compared with other proposed approximation quality measures according to the Proposition 15.

6. Attribute reduction in BRSM for multi-decision problems

Some approximation quality measures for multi-decision problems have been proposed in the former section. Though they have different semantic interpretation and can be described from the different profiles of a given decision table, it can be found that they share some common properties in nature. For any $\varphi_m \in \{megabs, metgabs, Pmeg, Nmeg, DPmeg, DNmeg\}$, an important theorem can be presented as follows:

Theorem 4. *Suppose a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $U/D = \{D_1, D_2, \ldots, D_{card(U/D)}\}$, $\forall B \subseteq C$, then $\varphi_m(D|B) \leq \varphi_m(D|C)$ and equality holds, if and only if $\forall D_j \in U/D$, it has $POS^*_C(D_j) \subseteq POS^*_B(D_j)$ and $NEG^*_C(D_j) \subseteq NEG^*_B(D_j)$.*

Theorem 4 indicates that each $\varphi_m \in \{megabs, metgabs, Pmeg, Nmeg, DPmeg, DNmeg\}$ has the property of monotonicity along with reducing attributes. If $\varphi_m(D|C) = \varphi_m(D|B)$, then the B-positive region and B-negative region of any decision class will not be reduced. Essentially, any quantitative measure φ_m discerns the objects between B-positive region and B-negative region for all decision classes. Based on these monotonic measures, an unified attribute reduction model can be constructed in Bayesian rough set model for multi-decision problems.

Given a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $\forall B \subseteq C$ if B satisfies the following two criteria:

(1) $\varphi_m(D|B) = \varphi_m(D|C)$;
(2) $\forall b \in B$, $\varphi_m(D|B - \{b\}) < \varphi_m(D|B)$.

Then B is called a reduct of C with respected to D, or called a reduct briefly.

For binary decision problems, $U/D = \{X, \neg X\}$, if $\varphi_m = metgabs$ and $\varphi_2 = \lambda$, according to $egabs(X|C) = egabs(\neg X|C)$, we have $\lambda(X|C) = \frac{metgabs(D|C)}{4P(X)P(\neg X)}$. Since $4P(X)P(\neg X)$ is a constant, attribute reduction model for binary decision problems can be considered as a special case of attribute reduction model for multi-decision problems.

In order to obtain reducts, the available heuristic attribute reduction algorithms in Pawlak's rough set model can also be used in BRSM. In the following, we will introduce the notion of discernibility matrix in BRSM for attribute reduction as dealing with multi-decision problems.

Given a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, its discernibility matrix can be defined as a $n \times n$ matrix $DM_m(DT) = (cm_{ij})_{n \times n}$, and the element cm_{ij} satisfies:

$$cm_{ij} = \begin{cases} \{a | a \in C \wedge \rho(x_i, a) \neq \rho(x_j, a)\} & \Omega \\ \emptyset & others \end{cases} \quad (37)$$

where Ω means $1 \leq j < i \leq n$ and $\exists D_k \in U/D$, such that $\delta_{D_k}(x_i)\delta_{D_k}(x_j) = -1$.

Obviously, it can be found that $DM_2(DT) = (c_{ij})_{n \times n}$ is a special case of $DM_m(DT) = (cm_{ij})_{n \times n}$.

Theorem 5. *Given a decision table $DT = (U, C \cup D, V, \rho)$ with multi-decision classes, $\forall B \subseteq C$, B is a reduct if and only if B satisfies the following two criteria:*

(1) $\forall i, j\ (1 \le j < i \le n)$, *if* $cm_{ij} \neq \emptyset$, *then* $B \cap cm_{ij} \neq \emptyset$;
(2) $\forall b \in B, \exists i, j, cm_{ij} \neq \emptyset$, *such that* $(B - \{b\}) \cap cm_{ij} = \emptyset$.

Proof. It can be proved as similar as Theorem 3. □

According to the Theorem 5, attribute reduction problems can be described by associated discernibility matrices in BRSM. All of the available attribute reduction algorithms based on discernibility matrices in Pawlak's rough set model can be analogously applied in BRSM. In addition, some notions in Pawlak's rough set model, such as attribute significance, core, discernibility function and others, can also be defined in BRSM both for binary decision and multi-decision problems, similarly.

7. Conclusions

BRSM can be applied to dealing with many practical problems, such as medical diagnosis, fault detection, economic forecasting and so on. Its main advantage is free of the user-defined parameters. In the model, the prior probability of the target event will be chosen as a benchmark value as defining the relevant approximation regions. In this paper, some properties and differences among BRSM, VPRSM and Pawlak's rough set model are analyzed. It can be proved that two kinds of current attribute reduction models in BRSM for binary decision problems are equivalent. Some monotonic approximation quality measures for multi-decision problems in BRSM are proposed and their relationships are also discussed. Since these monotonic measures share some common properties essentially, an unified attribute reduction model in BRSM is constructed. Therein, it is illustrated that attribute reduction model for binary decision problems is a special case of the one for muliti-decision problems.

Discernibility matrices for binary decision and multi-decision problems in BRSM, in which elements will discern the objects between the B-positive region and B-negative region of each decision class, are also discussed, respectively. So discernibility matrix based concepts and knowledge reduction approaches in Pawlak's rough set model can be analogously exploited in BRSM. All of the presented notions in this paper will be beneficial for the development of BRSM. The comparative investigation and experiments on rule sets in different probabilistic rough set models are our next work.

Appendix A

Proof of Theorem 2. Because $2P(X)(1 - P(X))$ is a constant after DT is given, so we only need to prove $egabs(X|B) \le egabs(X|C)$, and equality holds if and only if $POS_C^*(X) \subseteq POS_B^*(X)$ and $NEG_C^*(X) \subseteq NEG_B^*(X)$.

Since any $F \in U/B$ can be presented as $F = \cup\{E | E \in U/C, E \subseteq F\}$, then if $P(X|F) \ge P(X)$, we have:

$$\begin{aligned} P(F)\,(P(X|F) - P(X)) &= P(X \cap F) - P(X)P(F) = \sum_{E \subseteq F} (P(X \cap E) - P(X)P(E)) \le \sum_{E \subseteq F} |P(X \cap E) - P(X)P(E)| \\ &= \sum_{E \subseteq F} P(E)\,|P(X|E) - P(X)| \end{aligned} \tag{38}$$

and if $P(X|F) \le P(X)$, we also have:

$$P(F)\,(P(X) - P(X|F)) \le \sum_{E \subseteq F} P(E)\,|P(X|E) - P(X)|.$$

So $\sum_F P(F)|P(X|F) - P(X)| \le \sum_F \sum_{E \subseteq F} P(E)\,|P(X|E) - P(X)| = \sum_E P(E)\,|P(X|E) - P(X)|$

Consequently, $egabs(X|B) \le egabs(X|C)$.

Since $P(X|F) \ge P(X)$, equality in formula (38) holds, if and only if $P(X|E) \ge P(X)$ for all $E \subseteq F$. It indicates that equality in formula (38) holds if and only if $P(X|F) \ge P(X) \to \forall_{E \subseteq F} P(X|E) \ge P(X)$ and similarly, we get $P(X|F) \le P(X) \to \forall_{E \subseteq F} P(X|E) \le P(X)$. The first implication indicates that if $F \subseteq POS_B^*(X) \cup BND_B^*(X)$, then $E \subseteq POS_C^*(X) \cup BND_C^*(X)$ for all $E \subseteq F$ and the second implication indicates that if $F \subseteq NEG_B^*(X) \cup BND_B^*(X)$, then $E \subseteq NEG_C^*(X) \cup BND_C^*(X)$ for all $E \subseteq F$. Then we have $POS_B^*(X) \cup BND_B^*(X) \subseteq POS_C^*(X) \cup BND_C^*(X)$ and $NEG_B^*(X) \cup BND_B^*(X) \subseteq NEG_C^*(X) \cup BND_C^*(X)$.

It implies that only elementary sets in U/C being included in $POS_C^*(X) \cup BND_C^*(X)$ respectively or being included in $NEG_C^*(X) \cup BND_C^*(X)$ respectively can be merged under equivalence relation B. It is equivalent to $POS_C^*(X) \subseteq POS_B^*(X)$ and $NEG_C^*(X) \subseteq NEG_B^*(X)$.

Proof of Theorem 3. Denotes $U/C = \{E_1, E_2, \ldots, E_s\}$, and $U/B = \{F_1, F_2, \ldots, F_t\}$ respectively.

(1) Sufficiency: Without loss of generality, suppose $\exists F_k \in U/B$ can be presented as $F_k = E_p \cup E_q (p \neq q)$. According to formula (29), there is $x_i \in E_p$ and $x_j \in E_q$, such that $c_{ij} = \emptyset$, namely, $\delta_X(x_i)\delta_X(x_j) = 0$ or $\delta_X(x_i)\delta_X(x_j) = 1$. If not, $B \cap c_{ij} \neq \emptyset$, and then E_i and E_j can not be merged under attribute set B. According to Proposition 4, we have $POS_C^*(X) \subseteq POS_B^*(X)$ and $NEG_C^*(X) \subseteq NEG_B^*(X)$. So $\varphi_2(X|B) = \varphi_2(X|C)$.

$\forall b \in B, \exists i, j, c_{ij} \neq \emptyset$, such that $(B - \{b\}) \cap c_{ij} = \emptyset$, combined with condition (1) $B \cap c_{ij} \neq \emptyset$, we have $B \cap c_{ij} = \{b\} \neq \emptyset$. It indicates that $\rho(x_i, b) \neq \rho(x_j, b)$, $\delta_X(x_i)\delta_X(x_j) = -1$ and $\forall b' \in (B - \{b\})$, it has $\rho(x_i, b') = \rho(x_j, b')$. So $\exists E_p, E_q \in U/C(p \neq q)$, where $x_i \in E_p$ and $x_j \in E_q$, will be merged under attribute set $B - \{b\}$. According to the proofs of Theorem 2, it has $\varphi_2(X|B - \{b\}) < \varphi_2(X|B)$.

Therefore, it can be concluded that B is a reduct.

(2) Necessity: For $\forall E_p, E_q \in U/C(p \neq q)$, 1) if they can be merged under attribute set B, namely, $\exists F_k \in U/B$, such that $F_k = E_p \cup E_q(p \neq q)$. since B is a reduct, $\varphi_2(X|B) = \varphi_2(X|C)$. According to Theorems 1 and 2, $\forall x_i \in E_p$ and $\forall x_j \in E_q$, it has $\delta_X(x_i)\delta_X(x_j) \neq -1$, then $c_{ij} = \emptyset$; 2) if they won't be merged under attribute set B, then $\exists b \in B$, $\forall x_i \in E_p$ and $\forall x_j \in E_q$, such that $\rho(x_i, b) \neq \rho(x_j, b)$. In this case, if $\delta_X(x_i)\delta_X(x_j) \neq -1$, then $c_{ij} = \emptyset$; if $\delta_X(x_i)\delta_X(x_j) = -1$, then $c_{ij} = \{b | b \in B, \rho(x_i, b) \neq \rho(x_j, b)\} \neq \emptyset$. So $\forall i, j (1 \leq j < i \leq n)$, if $c_{ij} \neq \emptyset$, it has $B \cap c_{ij} \neq \emptyset$.

$\forall b \in B, \varphi_2(X|B - \{b\}) < \varphi_2(X|B) = \varphi_2(X|C)$, it indicates that $\exists E_p, E_q \in U/C(p \neq q)$ won't be merged under attribute set B, but can be merged under attribute set $B - \{b\}$, where $\forall x_i \in E_p, \forall x_j \in E_q$, it has $\delta_X(x_i)\delta_X(x_j) = -1$. So $\rho(x_i, b) \neq \rho(x_j, b)$ and $\forall b' \in (B - \{b\})$, it has $\rho(x_i, b') = \rho(x_j, b')$. According to formula (29), it has $b \in c_{ij}$ and $\forall b' \in (B - \{b\}), b' \notin c_{ij}$, thus $(B - \{b\}) \cap c_{ij} = \emptyset$. Namely, $\forall b \in B, \exists i, j, c_{ij} \neq \emptyset$, such that $(B - \{b\}) \cap c_{ij} = \emptyset$. □

Proof of Theorem 4

(1) $\varphi_m = megabs$.

Since any $F \in U/B$ can be represented as $F = \cup\{E | E \in U/C, E \subseteq F\}$, $\forall D_j \in U/D$, it has:

$$\begin{aligned} P(F)|P(D_j|F) - P(D_j)| &= |P(D_j \cap F) - P(D_j)P(F)| = \left|\sum_{E\subseteq F} (P(D_j \cap E) - P(D_j)P(E))\right| \leq \sum_{E\subseteq F} |P(D_j \cap E) - P(D_j)P(E)| \\ &= \sum_{E\subseteq F} P(E)\left|P(D_j|E) - P(D_j)\right|. \end{aligned}$$

So,

$$\begin{aligned} \sum_{F\in U/B} P(F)|P(D_j|F) - P(D_j)| &\leq \sum_{F\in U/B} \sum_{\substack{E\in U/C \\ \wedge E\subseteq F}} P(E)\left|P(D_j|E) - P(D_j)\right| \\ &= \sum_{E\in U/C} P(E)\left|P(D_j|E) - P(D_j)\right|. \end{aligned} \tag{39}$$

Thus, $P(D_j)\sum_{F\in U/B} P(F)|P(D_j|F) - P(D_j)| \leq P(D_j)\sum_{E\in U/C} P(E)\left|P(D_j|E) - P(D_j)\right|$ and $\sum_{D_j\in U/D} P(D_j)\sum_{F\in U/B} P(F)|P(D_j|F) - P(D_j)| \leq \sum_{D_j\in U/D} P(D_j)\sum_{E\in U/C} P(E)\left|P(D_j|E) - P(D_j)\right|$.

Consequently, $megabs(D|B) \leq megabs(D|C)$.

According to the proofs of Theorem 2, equality in formula (39) holds, if and only if $POS_C^*(D_j) \subseteq POS_B^*(D_j)$ and $NEG_C^*(D_j) \subseteq NEG_B^*(D_j)$.

So $megabs(D|B) = megabs(D|C)$ if and only if $\forall D_j \in U/D$, it has $POS_C^*(D_j) \subseteq POS_B^*(D_j)$ and $NEG_C^*(D_i) \subseteq NEG_B^*(D_i)$.

(2) $\varphi_m = metgabs$ can be proved as similar as (1).

(3) $\varphi_m = Pmeg$.

Since any $F \in U/B$ can be represented as $F = \cup\{E | E \in U/C, E \subseteq F\}$, $\forall D_j \in U/D$, if $F \subseteq POS_B^*(D_j) \cup BND_B^*(D_j)$, then:

$$\begin{aligned} P(F)\left(P(D_j|F) - P(D_j)\right) &= P(D_j \cap F) - P(D_j)P(F) = \sum_{\substack{E\in U/C \\ \wedge E\subseteq F}} \left(P(D_j \cap E) - P(D_j)P(E)\right) = \sum_{\substack{E\in U/C \\ \wedge E\subseteq F}} P(E)\left(P(D_j|E) - P(D_j)\right) \\ &\leq \sum_{\substack{E\in U/C\wedge E\subseteq F \\ \wedge E\subseteq POS_C^*(D_j)\cup BND_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right) = \sum_{\substack{E\in U/C\wedge E\subseteq F \\ \wedge E\subseteq POS_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right). \end{aligned} \tag{40}$$

So,

$$\begin{aligned} &\sum_{\substack{F\in U/B \\ \wedge F\subseteq POS_B^*(D_j)}} P(F)\left(P(D_j|F) - P(D_j)\right) = \sum_{\substack{F\in U/B \\ \wedge F\subseteq POS_B^*(D_j)\cup BND_B^*(D_j)}} P(F)\left(P(D_j|F) - P(D_j)\right) \leq \sum_{\substack{F\in U/B \\ \wedge F\subseteq POS_B^*(D_j)}} \\ &\sum_{\substack{E\in U/C\wedge E\subseteq F \\ \wedge E\subseteq POS_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right) \leq \sum_{F\in U/B} \sum_{\substack{E\in U/C\wedge E\subseteq F \\ \wedge E\subseteq POS_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right) = \sum_{\substack{E\in U/C \\ \wedge E\subseteq POS_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right). \end{aligned} \tag{41}$$

Thus $\sum_{D_j\in U/D} \sum_{\substack{F\in U/B \\ \wedge F\subseteq POS_B^*(D_j)}} P(F)\left(P(D_j|F) - P(D_j)\right) \leq \sum_{D_j\in U/D} \sum_{\substack{E\in U/C \\ \wedge E\subseteq POS_C^*(D_j)}} P(E)\left(P(D_j|E) - P(D_j)\right)$.

Consequently, $Pmeg(D|B) \leq Pmeg(D|C)$.

Where equality in formula (40) holds, if and only if $P(D_j|E) \geq P(D_j)$ is satisfied for all $E \subseteq F$. It indicates that if $F \subseteq POS_B^*(D_j) \cup BND_B^*(D_j)$, then $E \subseteq POS_C^*(D_j) \cup BND_C^*(D_j)$ for all $E \subseteq F$, namely, $POS_B^*(D_j) \cup BND_B^*(D_j) \subseteq POS_C^*(D_j) \cup BND_C^*(D_j)$.

In addition, equality in formula (41) holds, if and only if $\forall F \in U/B$, if $F \not\subset POS_B^*(D_j)$, then $E \not\subset POS_C^*(D_j)$ for all $E \subseteq F$. It means that $NEG_B^*(D_j) \cup BND_B^*(D_j) \subseteq NEG_C^*(D_j) \cup BND_C^*(D_j)$.

So, $Pmeg(D|B) = Pmeg(D|C)$ if and only if $\forall D_j \in U/D$, it has $POS_B^*(D_j)$ $\cup BND_B^*(D_j) \subseteq POS_C^*(D_j) \cup BND_C^*(D_j)$ and $NEG_B^*(D_j) \cup BND_B^*(D_j) \subseteq NEG_C^*(D_j) \cup BND_C^*(D_j)$. It is equivalent to $\forall D_j \in U/D$, $POS_C^*(D_j) \subseteq POS_B^*(D_j)$ and $NEG_C^*(D_j) \subseteq NEG_B^*(D_j)$.

(4) $\forall \varphi_m \in \{Nmeg, DPmeg, DNmeg\}$ can be proved as similar as (3).

Acknowledgements

The authors are grateful to the anonymous referees for their valuable comments and suggestions. This work was supported by National Natural Science Foundation of China (Serial No. 60475019, 60775036, 60970061) and Research Fund for the Doctoral Program of Higher Education in China (Serial No. 20060247039).

References

[1] Z. Pawlak, Rough sets, International Journal of Information and Computer Science 11 (5) (1982) 314–356.
[2] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[3] J. Jelonek, K. Krawiec, R. Slowinski, Rough set reduction of attributes and their domains for neural networks, International Journal of Computational Intelligence 11 (2) (1995) 339–347.
[4] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition 35 (11) (2002) 2425–2438.
[5] A.I. Dimitras, R. Slowinski, R. Susmaga, et al., Business failure prediction using rough sets, European Journal of Operational Research 114 (2) (1999) 263–280.
[6] C. Anantaram, G. Nagaraja, K.V. Nori, Verification of accuracy of rules in a rule based system, Data and Knowledge Engineering 27 (1998) 115–138.
[7] M.M. Mushrif, A.K. Ray, Color image segmentation: rough-set theoretic approach, Pattern Recognition Letters 29 (4) (2008) 483–493.
[8] Z. Pawlak, S.K.M. Wong, W. Ziarko, Rough sets: probabilistic versus deterministic approach, International Journal of Man-Machine Studies 29 (1) (1988) 81–95.
[9] M. Banerjee, K. Pals, Roughness of a fuzzy set, Information Sciences 93 (3–4) (1996) 235–246.
[10] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (2–3) (1999) 191–209.
[11] W. Zhu, F.Y. Wang, Reduction and axiomtization of covering generalized rough sets, Information Sciences 152 (2003) 217–230.
[12] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae 27 (1996) 245–253.
[13] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, European Journal of Operational Research 129 (2001) 1–47.
[14] K. Wojciech, D. Krzysztof, G. Salvatore, et al., Stochastic dominance-based rough set model for ordinal classification, Information Sciences 178 (21) (2008) 4019–4037.
[15] J.T. Yao, Y.Y. Yao, W. Ziarko, Probabilistic rough sets: approximations, decision-makings, and applications, International Journal of Approximate Reasoning 49 (2008) 253–254.
[16] S. Greco, B. Matarazzo, R. Slowinski, Parameterized rough set model using rough membership and Bayesian confirmation measures, International Journal of Approximate Reasoning 49 (2008) 285–300.
[17] W. Ziarko, Variable precision rough set model, Journal of Computer and System Sciences 46 (1) (1993) 39–59.
[18] J.K. Katzberg, W. Ziarko, Variable precision rough sets with asymmetric bounds, in: Proc. of the International Workshop on Rough Sets and Knowledge Discovery, October 12–15, Banff, Alberta, Canada, 1993, pp. 163–190.
[19] A. An, N. Shan, C. Chan, et al., Discovering rules for water demand prediction: an enhanced rough set approach, Engineering Application and Artificial Intelligence 9 (6) (1996) 645–653.
[20] G. Xie, J.L. Zhang, K.K. Lai, et al., Variable precision rough set for group decision-making: an application, International Journal of Approximate Reasoning 49 (2008) 331–343.
[21] M. Beynon, The elucidation of an iterative procedure to β-reduct selection in the variable precision rough set model, Proc. of RSCTC'2004, LNAI, vol. 3066, Springer-Verlag, 2004, pp. 412–417.
[22] Chao-Ton Su, Jyh-Hwa Hsu, Precision parameter in the variable precision rough sets model: an application, The International Journal of Management Science 34 (2) (2006) 149–157.
[23] M. Beynon, The introduction and utilization of (l, u)-graphs in the extended variable precision rough sets model, International Journal of Intelligent Systems 18 (10) (2003) 1035–1055.
[24] Z. Pawlak, New look on Bayes' theorem-The rough set outlook, in: S. Hirano, M. Inuiguchi, S. Tsumoto (Eds.), Proc. of RSTGC'2001, May 20–22, Matsue, Shimane, Japan, Bull. Int. Rough Set Society, vol. 5, 2001, pp. 1–8 (1).
[25] Z. Pawlak, A rough set view on Bayes' theorem, International Journal of Intelligent Systems 18 (5) (2003) 487–498.
[26] Y.Y. Yao, S.K.M. Wong, P. Lingras, A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova, M.L. Emrich (Eds.), Methodologies for Intelligent Systems, North-Holland, New York, 1990, pp. 17–24., 5.
[27] Y.Y. Yao, S.K.M. Wong, A decision theoretic framework for approximating concepts, International Journal of Man-machine Studies 37 (1992) 793–809.
[28] Y.Y. Yao, Decision-theoretic rough set modes, Proc. of RSKT'2007, LNAI, vol. 4481, Springer-Verlag, 2007, pp. 1–12.
[29] Y.Y. Yao, Probabilistic rough set approximations, International Journal of Approximate Reasoning 49 (2008) 255–271.
[30] Y.Y. Yao, Probabilistic approaches to rough sets, Expert Systems 20 (2003) 287–297.
[31] Y.Y. Yao, Yan Zhao, Attribute reduction in decision-theoretic rough set model, Information Sciences 178 (2008) 3356–3373.
[32] J.P. Herbert, J.T. Yao, Criteria for choosing a rough set model, Computers and Mathematics with Applications 57 (2009) 908–918.
[33] D. Slezak, W. Ziarko, Bayesian rough set model, in: Proc. of the International Workshop on Foundation of Data Mining(FDM'2002), December 9, Maebashi, Japan, 2002, pp. 131–135.
[34] D. Slezak, Attribute reduction in the Bayesian version of variable precision rough set model, Electronic Notes in Theoretical Computer Science 82 (4) (2003) 263–273.
[35] D. Slezak, W. Ziarko, Variable precision Bayesian rough set model, Proc. of RSFDGrC'2003, LNAI, vol. 2639, Springer-Verlag, 2003, pp. 312–315.
[36] T. Nishina, M. Nagamachi, H. Tanaka, Variable precision Bayesian rough set model and its application to Kansei engineering, Transactions on Rough Sets V, LNCS, vol. 4100, Springer-Verlag, 2006, pp. 190–206.
[37] T. Nishino, M. Nagamachi, H. Tanaka, Variable precision Bayesian rough set model and its application to human evaluation data, Proc. of RSFDGrC'2005, LNAI, vol. 3641, Springer-Verlag, 2005, pp. 294–303.
[38] S. Widz, K. Revett, D. Slezak, A hybrid approach to MR imaging segmentation using unsupervised clustering and approximate reducts, Proc. of RSFDGrC'2005, LNAI, vol. 3642, Springer-Verlag, 2005, pp. 372–382.

[39] D. Slezak, Rough sets and Bayes factor, Transactions on Rough Sets III, LNCS, vol. 3400, Springer-Verlag, 2005, pp. 202–229.
[40] D. Slezak, The rough Bayesian model for distributed decision systems, Proc. of RSCTC'2004, LNAI, vol. 3066, Springer-Verlag, 2004, pp. 384–393.
[41] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (3) (2001) 592–605.
[42] J.Y. Wang, J. Zhou, Research of reduct features in the variable precision rough set model, Neurocomputing 72 (10–12) (2009) 2643–2648.
[43] D. Slezak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximation Reasoning 40 (1–2) (2005) 81–91.
[44] W. Ziarko, Probabilistic approach to rough sets, International Journal of Approximate Reasoning 49 (2) (2008) 272–284.
[45] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Dordrecht, 1991, pp. 331–362.
[46] W. Ziarko, Set approximation quality measures in the variable precision rough set model, in: Proc. of the 2nd International Conference on Hybrid Intelligent Systems (HIS'02), Soft Computing Systems, vol. 87, 2002, pp. 442–452.
[47] W. Ziarko, Probabilistic rough sets, Proc. of RSFDGrC'2005, LNAI, vol. 3641, Springer-Verlag, 2005, pp. 283–293.
[48] W. Ziarko, Stochastic approach to rough set theory, Proc. of RSCTC'2006, LNAI, vol. 4259, Springer-Verlag, 2006, pp. 38–48.